

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-278579

(43)Date of publication of application : 27.09.2002

---

(51)Int.Cl. G10L 15/08

G10L 15/10

G06F 17/30

G10L 15/06

G10L 15/00

G10L 15/28

---

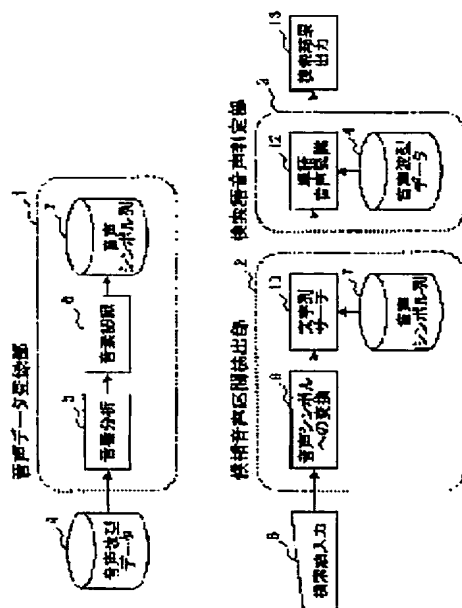
(21)Application number : 2001-077107 (71)Applicant : RICOH CO LTD

(22)Date of filing : 16.03.2001 (72)Inventor : KITAGAWA HIROO

---

(54) VOICE DATA RETRIEVING DEVICE

本発明の実施の形態に係る音声データ検索装置の構成図



(57)Abstract:

PROBLEM TO BE SOLVED: To provide a voice data retrieving device accurately searching a part, including a desired speech at a high speed from voice data stored in a large volume.

SOLUTION: This voice data retrieving device is constituted of a voice data registration part 1 converting digitized voice waveform data 4 to a preset voice symbol sequence and recording it, a candidate voice section detection part 2 for converting a retrieval word to a voice symbol and retrieving a matching part from a registered symbol sequence 7 and a retrieval word voice determining part 3 for determining whether or not a candidate section detected in

the candidate voice section detection part matches with the retrieval word. Thus, the candidate section is narrowed at a high speed at a symbol level and accurate detection by matching at a voice waveform level is conducted.

## LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

---

## CLAIMS

---

[Claim(s)]

[Claim 1] The voice-data retrieval equipment characterized by to have the voice-data registration section which changes and records the digitized voice data point on the voice symbol train which set up beforehand, the candidate voice section detecting element which search the part which changes a search term into a voice symbol and is in agreement out of a registration symbol sequence, and the search-term voice judging section which judge that the candidate section detected by this candidate voice section detecting element is in agreement with a search term.

[Claim 2] Voice data retrieval equipment characterized by for said voice data registration section memorizing the vocal parameter train used in case a voice symbol is extracted from voice data, and sharing this vocal parameter train in said search term voice judging section in voice data retrieval equipment according to claim 1.

[Claim 3] Voice data retrieval equipment with which said candidate voice section detecting element is characterized by using and searching a full-text search system to retrieval of a voice symbol in voice data retrieval equipment according to claim 1.

[Claim 4] Voice data retrieval equipment characterized by using for a judgment the word-spotting voice recognition unit with which said search term voice judging section is free with a voice recognition unit, and can treat the audio start edge and termination in voice data retrieval equipment according to claim 1.

[Claim 5] Voice data retrieval equipment characterized by said voice data registration section changing and using only a voiced vowel, a long vowel, a syllabic nasal, and the silent voice section for a voice symbol train in voice data retrieval equipment according to claim 1.

[Claim 6] Voice data retrieval equipment characterized by using the symbol group which carried out grouping of the consonant which is easy to start KONFUYUJON, and was made one syllable treatment as a voice symbol train in voice data retrieval equipment according to claim 1 although said voice data registration section is based on a single sound knot.

[Claim 7] voice data retrieval equipment according to claim 5 or 6 -- setting -- said voice data registration section -- the frequency of occurrence -- the voice data retrieval equipment which carries out dictionary registration of the high word beforehand, and is characterized by adding a word as one voice symbol.

---

---

## DETAILED DESCRIPTION

---

[Detailed Description of the Invention]

[0001]

[Field of the Invention] This invention relates to the voice data retrieval equipment with which the utterance part which includes the desired content from the voice data point accumulated in the large quantity is searched to accuracy at high speed.

[0002]

[Description of the Prior Art] Conventionally, the voice data retrieval equipment with which desired voice data is searched is proposed out of a voice data storage means by which voice data was memorized (JP,2000-020551,A). The word's by which carried out section division of voice data, and lexical selection was beforehand made in each voice section at time of registration existence probability is calculated and accumulated, and he decomposes into a retrieval word group including the synonym of the inputted search term, and is trying to output the voice section when those existence probabilities become the highest with this voice data retrieval equipment at the time of retrieval.

[0003]

[Problem(s) to be Solved by the Invention] However, by this method, in order for there to be a limit that only the search term by which dictionary registration was carried out beforehand can be treated and to ease this limit, when a vocabulary is increased, there is a problem of leading to large buildup of storage capacity. Furthermore, when the capacity of voice data increases, count of a retrieval word group's existence probability also has the problem that retrieval time will also increase in proportion to it, in order to carry out the full search of between whole tone vocal register.

[0004] then, such [ this invention ] a trouble -- taking an example -- \*\* -- the part which includes desired utterance from the voice data which is \*\* and was accumulated in the large quantity -- a high speed -- and it aims at offering the voice data retrieval equipment which can carry out a head broth to accuracy.

[0005]

[Means for Solving the Problem] The above-mentioned technical problem is solved by the means of the following this inventions.

[0006] The voice data retrieval equipment of invention according to claim 1 The voice data registration section which changes and records the digitized voice data point on the voice symbol train set up beforehand, It is characterized by having the candidate

voice section detecting element which searches the part which changes a search term into a voice symbol and is in agreement out of a registration symbol sequence, and the search term voice judging section which judges whether the candidate section detected by this candidate voice section detecting element is in agreement with a search term.

[0007] Moreover, in voice data retrieval equipment according to claim 1, said voice data registration section memorizes the vocal parameter train used in case a voice symbol is extracted from voice data, and the voice data retrieval equipment of invention according to claim 2 is characterized by sharing this vocal parameter train in said search term voice judging section.

[0008] Moreover, it is characterized by for said candidate voice section detecting element using a full-text search system at retrieval of a voice symbol, and the voice data retrieval equipment of invention according to claim 3 searching in voice data retrieval equipment according to claim 1.

[0009] Moreover, the voice data retrieval equipment of invention according to claim 4 is characterized by using for a judgment the word-spotting voice recognition unit with which said search term voice judging section is free with a voice recognition unit, and can treat the audio start edge and termination in voice data retrieval equipment according to claim 1.

[0010] Moreover, the voice data retrieval equipment of invention according to claim 5 is characterized by said voice data registration section changing and using only a voiced vowel, a long vowel, a syllabic nasal, and the silent voice section for a voice symbol train in voice data retrieval equipment according to claim 1.

[0011] Moreover, in voice data retrieval equipment according to claim 1, although said voice data registration section is based on a single sound knot for the voice data retrieval equipment of invention according to claim 6, it is characterized by using the symbol group which carried out grouping of the consonant which is easy to start KONFUYUJON, and was made one syllable treatment as a voice symbol train.

[0012] moreover, the voice data retrieval equipment of invention according to claim 7 -- voice data retrieval equipment according to claim 5 or 6 -- setting -- said voice data registration section -- the frequency of occurrence -- dictionary registration of the high word is carried out beforehand, and it is characterized by adding a word as one voice symbol.

[0013] In above-mentioned this invention, voice data is first changed into the symbol in which character expression is possible, and matching on this alphabetic character level extracts the candidate voice section. Next, pattern matching of a voice data

point is performed in each detected voice section, and it judges whether the search term is included actually. Rapidity and accuracy are realizable with two steps of this processing.

[0014]

[Embodiment of the Invention] Hereafter, the gestalt of operation of this invention is concretely explained based on a drawing.

[0015] Drawing 1 shows the block diagram of the voice data retrieval equipment concerning the gestalt of operation of this invention. As shown in drawing 1, voice data retrieval equipment consists of the voice data registration section 1, a candidate voice section detecting element 2, and the search term voice judging section 3. In the voice data registration section 1, the digitized voice data point 4 is changed and recorded on the voice symbol train 7 set up beforehand. The candidate voice section detecting element 2 searches the part which changes the inputted search term into a voice symbol, and matches out of the registration voice symbol sequence 7. a \*\*\*\*\*  
[ that the target search term is contained at the section about all the voice sections when the search term voice judging section 3 was detected by the alphabetic character search 10 ] -- the original voice data wave -- it judges using 4.

[0016] First, the voice data registration section 1 shown in drawing 1 is explained. The target voice data point 4 shall be digitized beforehand, and shall be stored in storage, such as a hard disk. The voice data point 4 is changed into characteristic quantity, such as spectrum information and power, for every time basis with short 5 - 10msec extent, and is outputted by sonagraphy 5 as a vocal parameter. In the phoneme recognition 4, the class of phoneme is specified from the time series data of this vocal parameter. Each extracted phoneme is mapped by the symbol group set up beforehand, doubles matching with the voice section, and accumulates it in a store (voice symbol train 7). Since the conversion precision to the voice symbol train 7 turns into detection precision at the time of retrieval, a voice symbol group does not express reading to accuracy, disregards the phoneme which is easy to mistake, or is treating some similar phonemes collectively.

[0017] Next, the candidate voice section detecting element 2 at the time of voice data retrieval is explained using drawing 1. A search term is inputted by input units, such as a keyboard, (search term input 8), and if the candidate voice section detecting element 6 is won popularity and passed, a search term is first mapped for the sequence of the symbol group of the same specification as having used at the time of voice data registration (conversion 9 as a voice symbol). A search term is the reading notation which expresses the content of utterance to look for and was written in

either a hiragana / katakana / Roman alphabet. It will become possible to treat the usual kanji kana mixture notation at the time of a search term input if a word dictionary is prepared separately.

[0018] In the character string search 10, the part which is in agreement with the voice symbol list of a search term out of the voice symbol train 7 registered into the voice data registration section 1 is extracted. If a coincidence part recognizes two or more existence, all the detection location of them will be outputted. The high-speed candidate section in symbol level narrows down by this, and it can \*\*.

[0019] Next, the search term voice judging section 3 at the time of voice data retrieval is explained using drawing 1 . In the search term voice judging section 3, it verifies whether the target search term is contained actually at the section using the original voice data point 4 about all the voice sections detected by the character string search 10 (word speech recognition 12). Although general word speech recognition specifies one out of two or more words registered, it inputs only a search term here, calculates the similarity, and he is trying to judge it with a threshold. Thereby, detection by the voice wave data level can be performed. When it is judged by the word speech recognition 12 that a search term exists, the location is outputted as a retrieval result (retrieval result output 13).

[0020] Next, the voice data retrieval equipment using a vocal parameter is explained using drawing 2 . Drawing 2 shows other block diagrams of the voice data retrieval equipment concerning the gestalt of operation of this invention. As shown in drawing 2 , the vocal parameter 26 obtained by processing of the sonagraphy 25 performed in the voice data registration section 21 is matched with a voice wave, and is memorized, and it is characterized by enabling it to use also for a search term voice judging in the search term voice judging section 23 at the time of retrieval. Although storage capacity increases a little, it can reduce the amount of operations at the time of retrieval, and becomes accelerable [ further ]. Since procedure is the same as the above-mentioned procedure, explanation is omitted here.

[0021] Next, the case where a full-text search system is introduced into the voice symbol retrieval in the character string search 9 at the time of retrieval is explained. As compared with searching the voice symbol train 11 in order flatly each time, high-speed search becomes possible about \*\*\*\*\* of the candidate voice section by having introduced the full-text search system into the voice symbol retrieval in the character string search 9 at the time of retrieval in the candidate voice section detecting element 2 shown in drawing 1 . Effectiveness is large when there are especially many amounts of voice data. However, in order that a full-text search

system may use a large-scale index file, the storage capacity required of a system increases.

[0022] Next, the case where a word-spotting speech recognition technique is used for the search term voice judging performed in the search term voice judging section 3 at the time of retrieval is explained. Here, word spotting is a recognition method which detects the target word, when the section is not limited to a voice pattern, but reference with a standard pattern is performed and coincidence likelihood looks for a high part.

[0023] By extending the ends of the candidate voice section a little, and passing a voice data point to word-spotting speech recognition by having used the word-spotting speech recognition technique in which it was free to the search term voice judging in drawing 1 performed in the search term voice judging section 3, and the audio start edge and termination could be treated to it, the incorrect recognition by omission of the phoneme of the initial of the word etc. can be mitigated, and a highly precise judgment is attained. Moreover, since word-spotting speech recognition needs to divide voice and does not need to give it, the judgment of whether the specified word exists is also equipped with it in the internal device, and it does not have to carry out threshold processing anew.

[0024] Next, the setting-out approach of the voice symbol group in the voice data registration section 1 is explained.

[0025] Here, all the consonants that are easy to cause incorrect recognition are disregarded, and only a voiced vowel, a long vowel, a syllabic nasal, and the silent voice section are changed and used for a voice symbol train. The error which the candidate section narrows down and is sometimes generated by this can be decreased, and a highly precise detection result is obtained. For example, supposing there is the utterance "audio retrieval ...", "on-EOENAUUIUE ..." which is the vowel sequence will be extracted as a voice symbol train. In this case, it is necessary to symbol-ize similarly the search term inputted in the search term input 8 at the time of retrieval. In the search term input 8 of drawing 1, when "ONSE" is inputted as a search term, a consonant is deleted from the pronunciation, and it changes into "ONE" and searches from the voice symbol train 11 registered (character string search 10). Here, all the words with the same vowel sequences, such as "respect", will be detected by the character string search 10 besides "voice." At the end, all the detected voice sections are checked by the word speech recognition 12, and only the target section is outputted as a retrieval result (retrieval result output 13).

[0026] Next, the case where what carried out grouping of the consonant which is easy



to carry out KONFUYUJON \*\* to the voice symbol group, and was considered as one syllable treatment is used is explained. Although based on a single sound knot, the candidate section narrowing down and stopping an error by carrying out grouping of the consonant which is easy to start KONFUYUJON, and using one symbol group which carried out syllable treatment as a voice symbol train 7, the cutback of the number of the candidate sections can be performed and more nearly high-speed retrieval is attained. for example, a silent burst -- when it is hard to identify "P" and "T" of a consonant, the syllable of a "PA" line and a "TA" line is treated as the same thing. About a notation, you may set it as freedom, such as assigning unification to a "PA" line and completely assigning unification and another symbol to a "TA" line. Since procedure is the same as the procedure mentioned above, explanation is omitted here.

[0027] Next, the case where a vowel or not only the short unit of a single sound knot but a word is assigned to a voice symbol group as one symbol is explained. the frequency of occurrence -- by carrying out dictionary registration of the high word beforehand, and assigning each word as one voice symbol, respectively, it also becomes reduction of the number of the candidate sections detected at the same time the voice symbol train 7 is reduced, and more nearly high-speed retrieval is attained. In order to extract an object word part at the time of registration of voice data, you may carry out from the detected phoneme sequence to the voice data registration section 1 in drawing 1 , and a word voice recognition unit may be independently prepared to it.

[0028] This invention is applicable to the system which searches and reproduces the part in which the argument on desired was made from the tape which recorded the board, the system which extracts the scene of a request of voice to a key from the video tape recorded on videotape.

[0029] Although the desirable example of this invention was explained in full detail above, various deformation and modification are possible for this invention within the limits of the summary of this invention which is not limited to the starting specific operation gestalt and was indicated by the claim.

[0030]

[Effect of the Invention] So that clearly from the place explained in full detail above invention according to claim 1 The voice data registration section which changes and records the digitized voice data point on the voice symbol train set up beforehand, Whether this candidate section is in agreement with a search term with constituting from the search term voice judging section to judge [ the candidate voice section

detecting element which searches the part which changes a search term into a voice symbol and matches out of a registration symbol sequence, and ] The exact detection by \*\*\*\*\* of the high-speed candidate section in symbol level and matching by the voice wave level is attained.

[0031] Moreover, in the voice data registration section, invention according to claim 2 memorizes the vocal parameter train used in case a voice symbol is extracted from voice data, by sharing this vocal parameter train in the search term voice judging section, can reduce the amount of operations at the time of retrieval, and becomes accelerable [ further ].

[0032] Moreover, in a candidate voice section detecting element, invention according to claim 3 is using a full-text search system for retrieval of a voice symbol, and although it accelerates \*\*\*\*\* of the candidate section further, it is possible.

Effectiveness is large in case especially the voice data of a large quantity is treated.

[0033] Moreover, incorrect recognition of word speech recognition can be decreased and a highly precise detection result is obtained because invention according to claim 4 uses for a judgment the word-spotting voice recognition unit which is free and can treat the audio start edge and termination in the search term voice judging section.

[0034] Moreover, invention according to claim 5 can disregard the consonant which is [ incorrect-] easy to recognize, the error which the candidate section narrows down and is sometimes generated by changing and using only a voiced vowel, a long vowel, a syllabic nasal, and the silent voice section for a voice symbol train can be decreased, and a highly precise detection result is obtained.

[0035] Moreover, although based on a single sound knot, invention according to claim 6 is carrying out grouping of the consonant which is easy to start KONFUYUJON, and using one symbol group which carried out syllable treatment as a voice symbol train, the candidate section narrowing it down and stopping an error, it can perform the cutback of the number of the candidate sections, and the more nearly high-speed retrieval of it is attained.

[0036] moreover, invention according to claim 7 -- the frequency of occurrence -- since dictionary registration of the high word is carried out beforehand and each word is assigned as one voice symbol, respectively, it also becomes reduction of the number of the candidate sections detected at the same time a voice symbol train is reduced, and more nearly high-speed retrieval is attained.

---

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号  
特開2002-278579  
(P2002-278579A)

(43)公開日 平成14年9月27日(2002.9.27)

(51)Int.Cl. <sup>7</sup>	識別記号	F I	テマコード <sup>*</sup> (参考)
G 1 0 L 15/08		G 0 6 F 17/30	1 7 0 E 5 B 0 7 5
15/10			3 2 0 C 5 D 0 1 5
G 0 6 F 17/30	1 7 0	G 1 0 L 3/00	5 3 1 W
	3 2 0		5 2 1 C
G 1 0 L 15/06			5 3 1 Z

審査請求 未請求 請求項の数7 O L (全 5 頁) 最終頁に続く

(21)出願番号 特願2001-77107(P2001-77107)

(22)出願日 平成13年3月16日(2001.3.16)

(71)出願人 000006747

株式会社リコー

東京都大田区中馬込1丁目3番6号

(72)発明者 北川 博雄

東京都大田区中馬込1丁目3番6号 株式  
会社リコー内

(74)代理人 100070150

弁理士 伊東 忠彦

Fターム(参考) 5B075 ND14 PP22 PR04

5D015 GG03 HH04

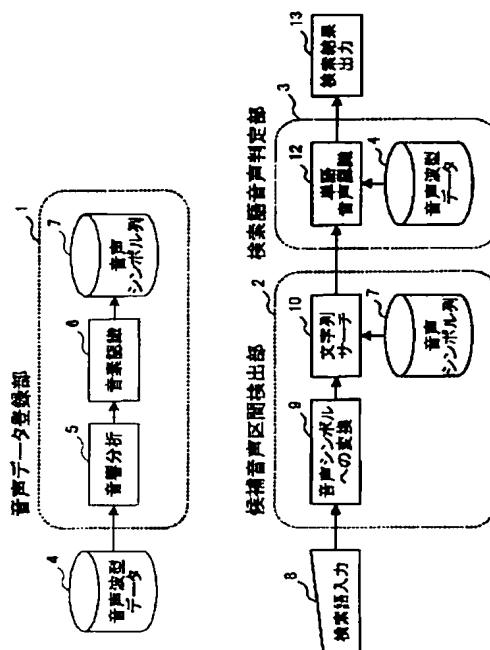
(54)【発明の名称】 音声データ検索装置

(57)【要約】

【課題】 大量に蓄積された音声データから所望の発話を含む部分を高速にかつ正確に頭だしできる音声データ検索装置を提供する。

【解決手段】 本発明の音声データ検索装置は、デジタル化された音声波形データ4を予め設定した音声シンボル列に変換して記録する音声データ登録部1と、検索語を音声シンボルに変換して登録シンボル系列7中から一致する部分を検索する候補音声区間検出部2と、候補音声区間検出部で検出された候補区間が検索語と一致しているか否かを判定する検索語音声判定部3とから構成されているので、シンボルレベルでの高速な候補区間の絞り込みと音声波形レベルでのマッチングによる正確な検出が可能となる。

本発明の実施の形態に係る音声データ検索装置の構成図

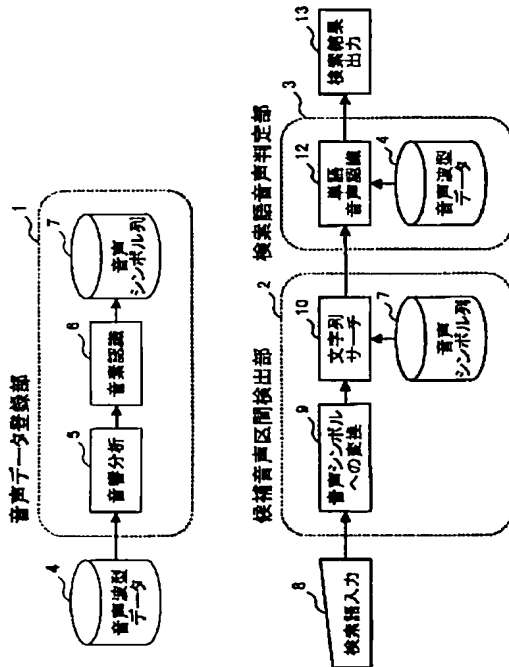


本とするが、コンフュージョンを起し易い子音をグループ化して1つの音節扱いしたシンボル群を音声シンボル列として用いることで、候補区間の絞込み誤りを抑えつつ、候補区間数の削減ができ、より高速な検索が可能となる。

【0036】また、請求項7記載の発明は、出現頻度高い単語を予め辞書登録しておき、各単語をそれぞれ1つの音声シンボルとして割り当てておくので、音声シンボル列が縮小されると同時に検出される候補区間数の減少にもなり、より高速な検索が可能となる。

【図1】

本発明の実施の形態に係る音声データ検索装置の構成図



【図面の簡単な説明】

【図1】本発明の実施の形態に係る音声データ検索装置の構成図である。

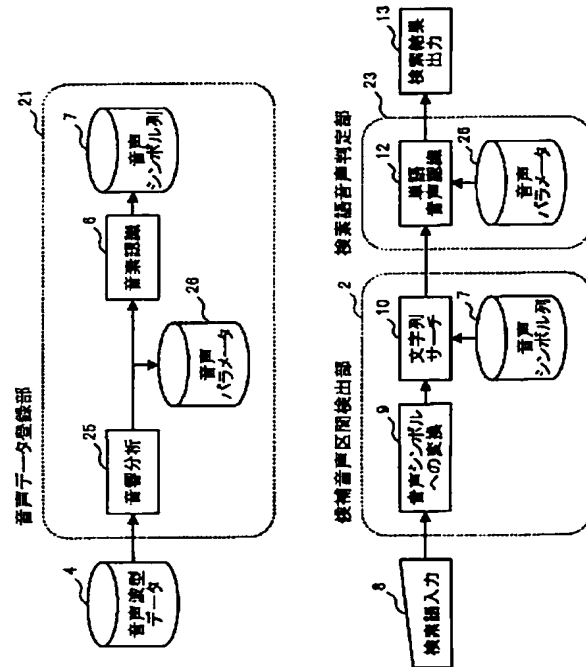
【図2】本発明の実施の形態に係る音声データ検索装置の他の構成図である。

【符号の説明】

- 1 音声データ登録部
- 2 候補音声区間検出部
- 3 検索語音声判定部

【図2】

本発明の実施の形態に係る音声データ検索装置の他の構成図



フロントページの続き

(51) Int. Cl.<sup>7</sup>

G10L 15/00  
15/28

識別記号

F I

G10L 3/00

5/06

ターコード (参考)

551P

571E

F